Title: A Machine Learns to Predict the Stability of Highly Diverse
Multi-planetary Systems

Author(s): Smullen, Rachel Ann
Ayyalapu, Neha

Intended for: Regeneron Science Talent Search 2022 (Science Fair)

Issued: 2021-11-05

A Machine Learns to Predict the Stability of Highly Diverse Multi-planetary Systems

Neha Ayyalapu

Abstract

Machine learning has proven to be an invaluable tool for characterizing the stability of planets in simplified planetary systems. In this work, we investigate the performance of a machine learning classifier on tightly-packed systems containing a rich diversity of planets, from Earths to Jupiters. Using information derived from short numerical simulations about a planet's early orbital evolution and its relationship with the most massive planets in the system, we train a random forest classifier to predict instability with a > 88 percent accuracy. Our classifier relies on relative planet masses and the standard deviation of eccentricity for much of its predictive power. Most misclassified planets lie along a multi-dimensional boundary between stable and unstable planets, indicating that their early orbital evolution is ambiguous. The major reason for misclassification in this work is timescale: because our classifier uses information from only the first 137 years of simulation data, it is blind to late time interactions that cause or prevent instability. Machine learning methods like those utilized in this work provide powerful tools to complement numerical simulations across a wide range of planetary architectures.

1 Introduction

Astronomers throughout history have speculated on the celestial dances of the planets. Yet, despite over 400 years of effort, understanding the evolution of planetary orbits remains an intricate, computationally expensive, and analytically unsolved problem.

1.1 Exoplanets

An exoplanet, or extrasolar planet, is a planet that lies beyond our Solar System. Since 1995, scientists have discovered more than 4,000 exoplanets with a variety of detection methods (Mayor & Queloz, 1995).

There have been five detection methods used to discover exoplanets: radial velocity, transit, direct imaging, gravitational microlensing, and astrometry. The transit and radial velocity methods are responsible for the vast majority of these discoveries. When a planet passes (or *transits*) between an observer and the central star it orbits, the planet blocks some of the light from the star. With the transit method, this minuscule decrease in brightness can suggest the presence of an exoplanet. The radial velocity method is dependent on the the gravitational force the planet exerts on the star: he planet's gravity causes the star to 'wobble' in space. The larger the planet, the more pronounced the resulting wobble is. When this phenomenon is observed with a telescope, it effects the the star's light spectrum. As the star moves in the direction of the observer, the color shifts towards blue; when moving away, it shifts towards red. This change in color is called redshift, and occurs because of the change in wavelength (of visible light waves) as the planet orbits around the star. By observing these changes in the color of light, astronomers are able to detect the presence of a planet.

Besides their use for discovering and validating the existence of exoplanets, these methods (especially when used together) offer vital information about both an individual planet and the exoplanetary system as a whole. They can inform the planet's mass, density, size, and distance from central star. With this, astronomers are often able to make determinations about a planet's composition, structure, and atmospheric composition.

Exoplanet discovery missions such as *TESS* and *Kepler* have revealed a vast diversity of exoplanets in our galaxy. Astronomers have seen compositions ranging from rocky (like Earth) to gas-rich (such as Jupiter). There have been planets dominated by water or ice (Neptune) and others by carbon or iron (like Mercury).

While these worlds are made of elements similar to those present in our Solar System, we have yet to find an entire system similar to our own. Understanding the architectures of real exoplanetary systems–and how that might relate to our own Solar System–requires us to understand the dynamics of diverse exoplanetary systems.

1.2 Planetary Dynamics

Orbits in space (the paths that bodies travel as they move around one another under the influence of gravity) are characterized by six Keplerian orbital elements—semi-major axis, eccentricity, inclination, argument of pericenter, longitude of ascending node, and true anomaly—which describe the time-dependent position of one body around another. Semi-major axis is a measure of the average extent of an orbit. The eccentricity is the ellipticity of the orbit, where $e = 0$ is a circle. Inclination measures the tilt of an orbit relative to a reference plane. The argument of pericenter is the location of the closest approach of an orbit relative to the point it crosses the reference plane. Longitude of ascending node is the angle between the reference direction and the point at which the orbit crosses from under the reference plane to over it, and true anomaly is the position of the planet at the current time.

The motions of bodies in the sky have been of interest since the ancient Greeks used circular orbits to describe the motion of the planets, the Moon, and the Sun. Centuries after the Ptolemaic model, in the early 1600s, Johannes Kepler developed the modern theory of orbits and the three laws of planetary motion. Orbits are, arguably, even more critical to understand today than they were hundreds of years ago. They are the easiest property to measure for exoplanetary systems, and, as an intrinsic component of our understanding of planets, they can help us understand the formation and long-term evolution of planets.

Decades after Kepler's laws of planetary motion, Isaac Newton proposed his law of universal gravitation, which describes the effects of gravity between two bodies. Combined, Kepler and Newton's work revolutionized celestial mechanics and left an open question about the long-term stability of our Solar System: do small gravitational perturbations between planets settle over large timescales or accumulate until a planet is rendered unstable?

Signatures of Dynamical Instability

Dynamical instability has been defined in several different ways. Some works assume that a system is experiencing instability when planet orbits come within one Hill radius of each other; others assume that instability occurs when two orbits cross. For this work, I assume that instability manifests as collisions

(either planet-planet or planet-star) or ejections from the planetary system.

Any system of three or more planets has no formally predictable solution. However, there are cases in which we can estimate the long-term orbital stability of a system. The simplest case is the Hill stability of two planets (masses m_1 and m_2) on circular orbits (at distances a_1 and a_2) around a star of mass M. If we define the spacing for stable orbits in terms of the mutual Hill radius ($R_{H,m}$), where $R_{H,m}$ is defined as

$$R_{H,m} = \left(\frac{m_1 + m_2}{3M}\right)^{1/3} \frac{a_1 + a_2}{2}, \tag{1.1}$$

we can then define the dimensionless dynamical spacing Δ of the two planets as

$$\Delta = \frac{a_2 - a_1}{R_{H,m}}. \tag{1.2}$$

Gladman (1993) calculates that two planets will be stable for all time as long as they satisfy $\Delta > 2\sqrt{3}$. For systems of three or more planets, it has been found that planets must be separated by $\Delta \gtrsim 10$ for Gyr stability (Chambers et al., 1996).

Hill stability is not the only metric to predict stability. Mean motion resonance overlap is found to lead to instability when the following condition is met (Wisdom, 1980; Deck et al., 2013)

$$\left(\frac{a_2 - a_1}{a_1}\right) \lesssim 1.46 \left(\frac{m_1 + m_2}{M}\right)^{2/7}. \tag{1.3}$$

Other works rely on short numerical simulations instead of initial conditions to predict a system's instability. For instance, chaos indicators such as MENGO (Mean Exponential Growth factor of Nearby Orbits) have been found to correlate with instability (Cincotta et al., 2003). No method has been able to theoretically predict the infinite stability of all planetary systems, and it is likely that none ever will. However, each metric provides some insight into the probable future evolution of a planet.

Numerical Simulations

Research in the past few decades has leveraged numerical N-body simulations to understand the role of dynamical evolution in shaping observed populations of exoplanets. For instance, dynamical scattering has been proposed to sculpt the observed distributions of planetary orbits and masses. Works such as Chambers et al. (1996), Faber & Quillen (2007), Smith & Lissauer (2009), Shikita et al. (2010), Pu & Wu (2015), Obertas et al. (2017), Wu et al. (2019), Gratia & Lissauer (2021) and Lissauer & Gavino (2021), among others, have found the relationship that increased dynamical planet spacing (discussed further in Section 1.2) leads

to increased stability times. Other works including Jurić & Tremaine (2008), Chatterjee et al. (2008), and Raymond et al. (2010) find that dynamical scattering leads to broader eccentricity distributions of planetary orbits that may match some subset of exoplanet observations. Mean motion resonances (and mean motion resonance overlap) also play a substantial role in planetary stability. Resonance overlap leads to shorter instability timescales as the forcing on a planet is much stronger and more chaotic (e.g., Chambers et al., 1996; Lissauer et al., 2011a; Morrison & Kratter, 2016; Obertas et al., 2017; Wu et al., 2019; Lissauer & Gavino, 2021). Numerical simulations have also been used to infer if planetary systems form with too many planets (overpacked; susceptible to instability), if they form with so many planets that moving one would introduce instability (maximally packed), if they form with enough planets that adding one more would cause instability (minimally packed), or if they form with few enough planets that more could be introduced without issue (sparse) (e.g., Barnes & Quinn, 2004; Fang & Margot, 2013; Kratter & Shannon, 2014; Agnew et al., 2019).

The major drawback to these types of numerical simulations of planetary interactions is that the problem is computationally challenging when studying the entire parameter space. Many of the aforementioned works therefore reduce the complexity of the problem and study systems with a limited number of planets (typically three to five), or only equal mass planets, or planets with equal dynamical spacing. This causes ambiguity in the applicability of these results directly to observed exoplanetary systems, which are never as idealized as simulations.

1.3   Machine Learning Applications

With the diversity of planetary systems that have been discovered and characterized to date, it becomes critical to understand the dynamical interactions and evolution of planets so that we can better infer both the processes that form these systems and their future evolution.

Machine learning can help broaden the parameter space in which we can understand detailed scattering outcomes by reducing computational cost of individual simulations and by introducing high-dimensional sets of initial conditions in simulations. Tamayo et al. (2016) was among the first works to combine machine learning and numerical simulation for predicting instability; they predicted stability of planets in planetary systems containing three $5\,M_\oplus$ planets with an average accuracy $\gtrsim 90$ percent. Since then, machine learning methods have been used to great effect in orbital dynamics. The applications have been used to predict the stability of circumbinary planets (Lam & Kipping, 2018), classify asteroid families (Carruba et al., 2020), classify Kuiper belt objects (Smullen & Volk, 2020), predict long-term instability in diverse three planet

systems (Tamayo et al., 2020), and to directly predict the instability time of three planet systems (Cranmer et al., 2021), among others.

In this paper, I present a machine learning classifier trained on the very early timescales of highly diverse, overpacked systems of 10 planets that has learned to predict the instability of individual planets. I provide an overview of my simulations, my creation of dynamically-motivated features upon which to train the classifier, and the training of the classifier itself in Section 2. I then examine the performance of the classifier, including which features are most indicative of stability, in Section 3. Finally, in Section 4, I discuss the reasons that my classifier can fail and compare it to similar methods from the literature.

2 Methods

In this section, I describe the methods used in this paper. I begin with numerical simulations of planetary systems and then derive metrics from those simulations that are then used in my machine learning method. Finally, I give a description of the random forest classifier method used to predict stability and the optimization thereof.

2.1 Simulations

The planetary systems used in this work come from the 'Mordaisini' single-star ensemble from Smullen et al. (2016). Ten planets were placed around a 1 $M_\odot$ star and integrated to 10 Myr using the Gauss-Radau variable timestep integrator in the MERCURY6 N-body integration package (Chambers & Migliorini, 1997). The initial orbital semi-major axis (a) and masses for each of the ten planets in each system were drawn from the distributions calculated in the population synthesis models from Mordasini et al. (2009a,b) as seen in Figure 2.1.

The semi-major axes span a range from about 0.1 to 15 a.u. with a peak at 3 a.u. (note that, because the comparison in Smullen et al. 2016 was to circumbinary planets, I do not have a population of short period inner planets, e.g., Lissauer et al., 2011b). Planet masses span $1 - 10^4 \, M_\oplus$; there is a dominant peak at low mass ($\sim 1 - 2 \, M_\oplus$) and small peaks around 1 Neptune mass and 1 Jupiter mass. Eccentricity and inclination are drawn from Rayleigh distributions with scale parameters of $e = 0.1$ and $i = 0.73°$. The remaining orbital elements, argument of pericenter and angle of the ascending node, are drawn from a uniform distribution. Planets are allowed to eject from the system (when they reach a distance of more than 1000 au from the star), collide with the central star, or collide with another planet. The system state is output every 5000 days
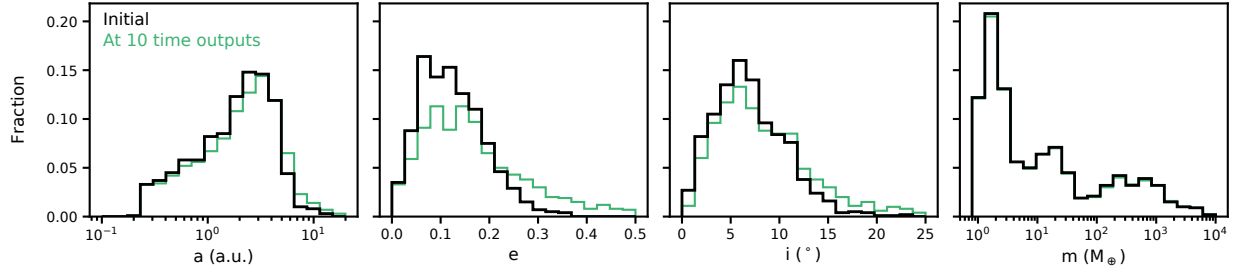
Figure 2.1: Orbital properties for the planet population used in this work. From left to right, the panels show semi-major axis, eccentricity, inclination, and mass. The black line shows the initial conditions, while the green line shows the orbital elements at 10 time outputs ($\approx 137\,\mathrm{yr}$), which is the averaging window used for computing features. Planets quickly begin to experience scattering events with other planets in the systems, as seen by the rapid broadening of the orbital element distributions.

($\approx 13.7\,\mathrm{yr}$).

These planetary systems are designed to be overpacked; as such, a large amount of orbital evolution and instability is expected. Planetary systems in this sample finish the 10 Myr simulations with an average of three planets per system, although individual system multiplicities range from 1–8 planets. About 70 percent of unstable planets are lost in the first 10^5 yr. Nearly half of all planets are lost via ejections, while about twice as many of the remaining unstable planets suffer collisions with other planets ($\sim$14 percent of all planets) than collisions with the central star ($\sim$8 percent of all planets). For the purpose of this work, I consider a planet to be *stable* if it lasts to 10 Myr and *unstable* if it otherwise ejects or collides before 10 Myr.

2.2 Derived Features

Machine learning methods utilize *features* of the data, which, for the random forest classification method used herein, are a set of reduced quantities derived from the simulations. In this section, I describe the choice and derivation of the twelve features that are used in this work to classify orbital instability.

Smullen et al. (2016) notes that the most massive planet in a system drives the final system multiplicity in these highly diverse planetary systems. Therefore, I compute features with the idea that the three most massive planets in the system likely have the largest influence on the dynamical outcomes of other planets. The planets in my systems span a large range in masses and semi-major axis, so I reduce bias in the set of features by ensuring all features are dimensionless. In addition to considering the osculating orbital elements, I derive mass ratio, semi-major axis ratio, and spacing in mutual Hill radii (equation 1.2) with the three most massive planets.

Table 2.1: Features used for classification

Symbolic Description	Explanation
Initial m/m_1	Initial mass ratio with most massive planet
Initial m/m_2	Initial mass ratio with 2nd most massive planet
Initial m/m_3	Initial mass ratio with 3rd most massive planet
$\max(e)$	Maximum e in time
$\sigma(e)$	Std. dev. of e in time
$\sigma(a/a_1)$	Std. dev. of a ratio with most massive planet in time
$\sigma(a/a_2)$	Std. dev. of a ratio with 2nd most massive planet in time
$\sigma(a/a_3)$	Std. dev. of a ratio with 3rd most massive planet in time
$\sigma(\Delta_{m_1})$	Std. dev. of Δ with most massive planet in time
$\sigma(\Delta_{m_2})$	Std. dev. of Δ with 2nd most massive planet in time
$\sigma(\Delta_{m_3})$	Std. dev. of Δ with 3rd most massive planet in time
$\mathrm{CV}(i)$	Coefficient of variation (mean/std. dev) of i in time

To best leverage information from the early orbital histories of planets in the simulations, I then create features by combining information over several time outputs. The simulations record the six orbital elements for each individual planet. Independently, these quantities offer minimal information about the future interactions between planets, so it is useful to compute dynamically-motivated quantities for each planet in the selected time domain.

I explore a wide set of potential features, including initial and final values, and the minimum, maximum, mean, standard deviation, and coefficient of variation of values in time. Each of these may encode information which could be significant in predicting orbital stability. I investigate classifier performance with several different time averages spanning 3 to 100 time outputs. I find that 10 time outputs (137 years) is the best optimization of time and performance; 137 years is the fiducial time average used throughout the rest of this paper. This time span equates to approximately 4300 orbits of an inner planet, 25 orbits of an average planet, and two orbits of an outer planet.

An excessive number of features can make the classifier output more difficult to interpret and can cause overfitting, which then impacts the accuracy and usability of my model. To reduce the ¿100 potential features to a smaller set, I use metrics of my classifier's feature importance (how significantly the computed features contribute to the classification) and the feature correlation coefficients to inform the final feature set. The matrix of correlation coefficients instructs us on the orthogonality between any two features; my final feature list aims to minimize correlated variables to avoid 'double counting' features by giving excess statistical weight with highly correlated quantities. The final set of features that are both significant to the classifier and minimally correlated are listed in Table 2.1.

2.3   Machine Learning

Machine learning classification is a family of methods that sort objects into different classes (categories) based on combinations of their features (the data attributed to each object). Broadly, machine learning is based on the idea that machines can learn from data, identify patterns, and make classifications on new data using these patterns.

I select the gradient boosting random forest classifier method, which is a supervised machine learning algorithm. Versions of this algorithm have been utilized in previous works such as Tamayo et al. (2016), Tamayo et al. (2020), and Smullen & Volk (2020). As a supervised algorithm, it first trains on data where it knows the correct classes–in this case, stable or unstable– and can then be used to predict the classes of unseen data. A random forest classifier creates an ensemble of decision trees, a flowchart-like structure with binary nodes representing specific features in the data. It then averages the outcome of each tree to determine the classification.

To create a machine learning model that can make predictions on data, I use a training-testing scheme. The classifier *trains* on a fraction of the data set to learn the features of specific classes. It then *tests* on the remaining data to estimate how well the model fits. When I begin the training process, an arbitrary 'line' is drawn through the multiple dimensions of the feature data. As training progresses, this line moves closer to the correct separation of the classes. Testing then tells us how well the model performs on unseen data. A standard testing-training split is 70 percent training data and 30 percent testing data, which I adopt here (my results are statistically similar with training fractions higher than 70 percent). Training/testing sets are typically split on a random sample of all objects. However, to preserve a valid testing set on an entire planetary system, I split the sample randomly by system. Table 2.2 shows a breakdown of the planetary data used in both the training and testing sets, with a breakdown by planet fate.

I refine the hyperparameters (the variables that tell the classifier how to behave) to make small improvements to the accuracy of the classifier. To do so, I use a 5-fold cross validation grid search. The data is first divided into five 'folds' (randomly split equal fractions of the data). The classifier is then trained on four of five folds and tested on the fifth. In each of the five iterations per set of hyperparameters, a different fold is used as the testing set. The final result is the average of score between the five folds. A grid search coupled with the cross validation then allows us to optimize the best hyperparameters for a specific algorithm and set of features.

I optimize for four hyperparameters: learning rate, maximum tree depth, maximum features, and number of estimators. The learning rate controls how quickly the classifier adjust the weights of the parameters.

Table 2.2: Data used in classifier. Columns show the total number of planets in each sample and the counts of stable planets, unstable planets, ejected planets, planets suffering collisions with the star, and planets colliding with other planets

	N	N_{stable}	$N_{unstable}$	N_{ej}	N_{s-col}	N_{p-col}
Testing	700	198	502	344	106	52
Training	300	88	212	144	37	31
Total	1000	286	714	488	143	83

Maximum tree depth is the number of vertical nodes in a given tree; setting a maximum depth limits complexity and prevents overfitting. The maximum features parameter controls the size of the feature subsets to consider when splitting at a node. Finally, a Random Forest is comprised of several individual trees, and the number of estimators controls the number of trees used in the forest. The optimal solution I arrive at has a learning rate of 0.025, a maximum tree depth of 3, a maximum number of features scaled by the square root of the number of samples, and 130 estimators.

Final Classifier Performance

Upon refining the hyperparameters, my classifier achieves an overall accuracy of 88.33 percent; this accuracy determines how many items I classified correctly out of all items I classified.

While accuracy is a common measurement of correctness, there are several other metrics useful to evaluate the performance of a machine learning classifier. Precision, for instance is a measure of how many correct samples a classifier returns out of all the items it returns in a class; it is a measure of how valid the results are. Recall, is a measure of how many positive cases the classifier correctly predicted; it is a measure of how complete or sensitive my classifier is.

For stable planets, precision is 86.3 percent and recall is 71.6 percent. For unstable planets, precision is 89.0 percent and the recall is 95.3 percent. This indicates that my classifier is notably more sensitive to signs instability than it is to signs of long-term stability, which is unsurprising given the quantity of unstable vs. stable planets in the planet population. However, I am still able to achieve relatively high accuracy on all types of planets (further discussed in Section 4.1).

3 Results

In this section, I describe the characteristics of my machine learning classifier and its performance on planets of different properties. Throughout the rest of this paper, I will use the following terms for brevity: *CSS*
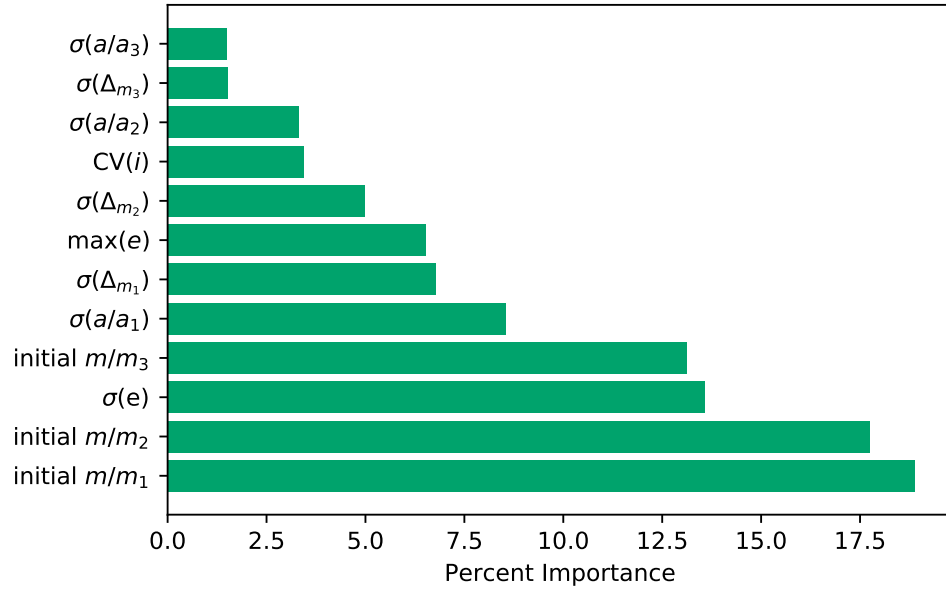
Figure 3.1: Feature importances for the twelve features used in my fiducial classifier. The sum of all feature importances adds up to 100 percent. Mass ratios with the most massive planets in the system and the planet eccentricities are the most important features for an accurate classification of unstable planets in this planetary population.

indicates planets that were correctly classified as stable, *CUU* indicates planets that were correctly classified as unstable, *MSU* indicates planets that were misclassified as stable but have a true classification of unstable, and *MUS* indicates planets that were misclassified as unstable but were stable to 10 Myr in the simulations.

3.1 Feature Importance and Classification Probability

Although I initially tested a wide range of features, I find the relatively small set of dimensionless features shown in Figure 3.1 allow the classifier to perform with an accuracy above 88 percent. Features used by a random forest classifier can be ranked in order of their importance to the classifier, where the sum of all feature importances adds up to 100 percent. My fiducial classifier relies most upon the mass ratios and eccentricities of a planet.

The initial mass ratio with the first, second, and third most massive planets in the entire planetary system are the first, second, and fourth most important features, respectively. These three features alone account for more than 45 percent of the total predictive power of the classifier. In my model, the mass ratio with the most massive planet in the system is the most important feature, followed very closely by the mass ratio with the second most massive planet in the system. The significance of these two features comes in part from the structure of the initial planetary systems. About 14 percent of all planets in the simulations are Jupiter mass planets or larger, and 22 percent are Saturn mass or larger. Therefore, there are, on average, two gas giants in each planetary system. These gas giants will dominate the orbital evolution of nearby smaller planets, with the impact being more extreme with a larger mass difference.
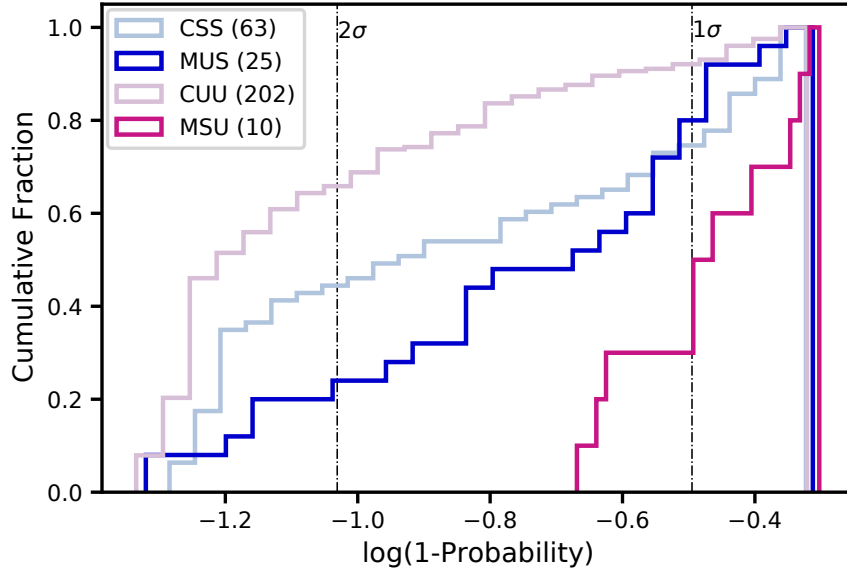
Figure 3.2: The cumulative probability distribution for correctly classified planets CSS (light blue) and CUU (light purple), and for misclassified planets MUS (dark blue) and MSU (dark purple). From right to left, the dashed lines represent 1σ (68%) and 2σ (95.4%). The correctly classified planets have notably higher class membership probabilities.

As opposed to the mass ratio, which may be thought of as being related to the *potential* for orbital instability, the variations in eccentricity (which are manifested in my feature list as $\sigma(e)$ and $\max(e)$) and semi-major axis (standard deviation of semi-major axis ratios with the massive planets) are related to *dynamic* scattering events early in the simulations. Large variations in eccentricity arise from orbital evolution and frequently predict later instability. The standard deviation of eccentricity–the third most important feature–and maximum eccentricity account for about 20 percent of the total predictive power of my classifier. The standard deviation of the semi-major axis ratio with the most massive planet (the planet typically doing the strongest scattering) accounts for over 8 percent of the classifier's predictive power.

The remaining feature importances include information about additional semi-major axis ratios, spacing in mutual Hill radii, and inclination. While these features are not individually very significant, their exclusion from the classifier leads to a substantially worse result. All of these additional features include some measure of the time variability of orbital evolution: the standard deviation of semi-major axis ratios measures how the planet separations are physically changing, the standard deviation of spacing in mutual Hill radii indicates how the dynamical planet separations are changing, and the coefficient of variation of inclination is an indication of scattering.

In addition to knowing what information the classifier needs to output an overall accurate classification, it is useful to know how reliable the classifications are for individual planets. To this end, I show the probability distributions of the four types of classifications (CSS, CUU, MUS, and MSU) in Figure 3.2. Over 60 percent of CUU planets have a probability of class membership greater than $2\sigma (\approx 95$ percent$)$, and about 40 percent of CSS classifications are above 2σ. Conversely, less than 20 percent of unstable planets
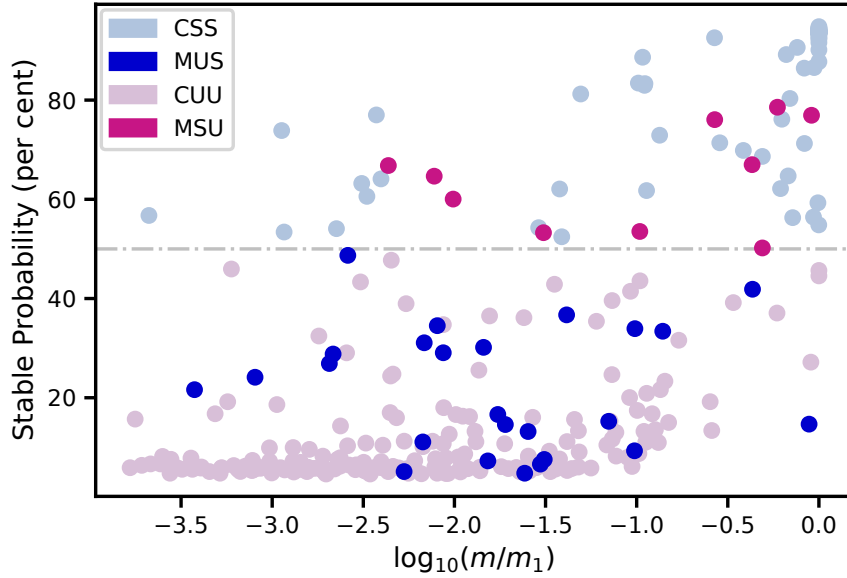
Figure 3.3: Stable class membership probability as a function of our most important feature, mass ratio with most massive planet. The dashed line at 50% represents the threshold for stable (above the line) vs. unstable (below the line) classification. The coloring convention is the same as that in Figure 3.2.

have a high probability of belonging to their identified class. Both of the misclassified curves rise steeply at low probabilities. About $20 - 40$ percent of MUS planets have relatively high unstable classification probabilities: these are planets that undergo initial scattering in the simulation but then finish the simulation on a stable orbit. The reasons for misclassification will be examined further in Section 4.1.

3.2 Classifier Performance and Planet Properties

Classification depends on a high dimensional combination of features, but I can gain some intuition about how the classifier works by looking at individual features of planets. Figure 3.3 shows the probability of stable classification plotted against the most important feature, the mass ratio with the most massive planet. Typically, stable planets have higher mass ratios and unstable planets have lower mass ratios, which follows from a simple analysis of the scattering outcome of two planets: the more massive planet will have less of an impacted orbit due to equal but opposite forces. Most of the misclassified planets trend to different masses than expected; MUS planets generally exhibit lower mass ratios, while the MSU planets trend toward higher mass ratios. My classifier assigns the highest probabilities to the extremes: very massive planets typically have a very high (> 80 percent) probability of a stable classification, while the lowest mass planets typically have a high probability of an unstable classification. Planets with mid-range mass ratios tend to have somewhat lower probabilities, suggesting that the classifier finds their combination of features more ambiguous.

Figure 3.4 shows the parameter space covered by the first and third most important features; here, the
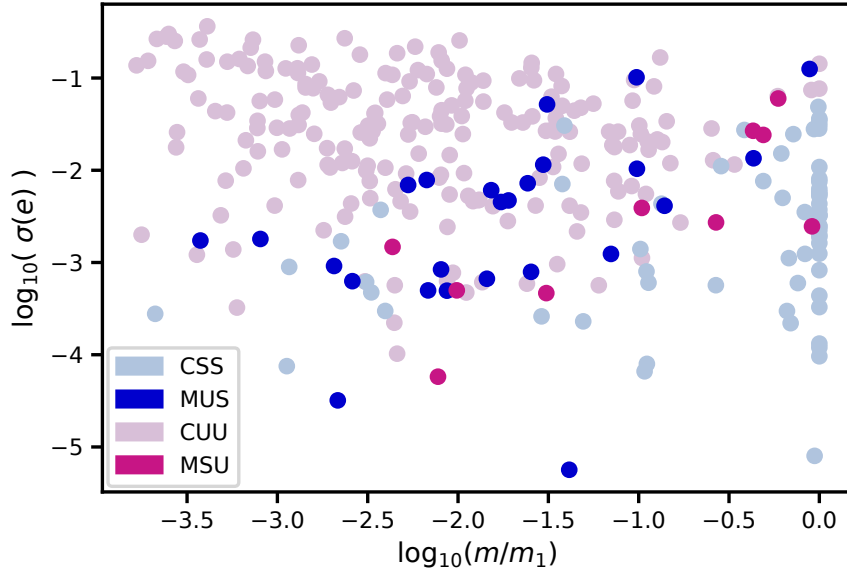
Figure 3.4: Standard deviation of eccentricity vs. mass ratio with most massive planet. The coloring convention is the same as that in Figure 3.2. Misclassified planets tend to lie along the boundary between the stable and unstable populations.

'thought process' of the classifier becomes even more apparent. All of the misclassified planets lie along the manifold separating the stable and unstable planets, or they lie in a space not populated by many other planets. This again demonstrates that my classifier is very good at identifying obviously stable and unstable planets but may encounter ambiguity in the less common boundary cases.

3.3   Classifier Performance and System Multiplicity

The fiducial classifier in this work, while only considering individual planets, has features derived from the entire planetary system. Figure 3.5 shows the classification accuracy as a function of system multiplicity. I find very high overall accuracy in planet classification for low multiplicity systems (1–3 planets remaining at 10 Myr): the majority of low multiplicity systems have more than 80 percent of all planets correctly classified. These are representative of the most common systems in the simulation ensemble. As the final multiplicity increases, my overall classification accuracy in a system decreases. Even at relatively high multiplicity, however, I am able to accurately classify unstable planets with an accuracy greater than 75 percent. My classifier sees the worst performance with stable planets, an unsurprising trend given these types of planets are both less represented in the training set (leading to less statistical significance from which the classifier can learn) and that these planets can still undergo scattering events that resemble catastrophic instability on short timescales.

The highest multiplicity systems (one with 8 planets remaining at 10 Myr and two with 6 planets remaining) consistently have the poorest performance across all classifiers tested in the course of this work.
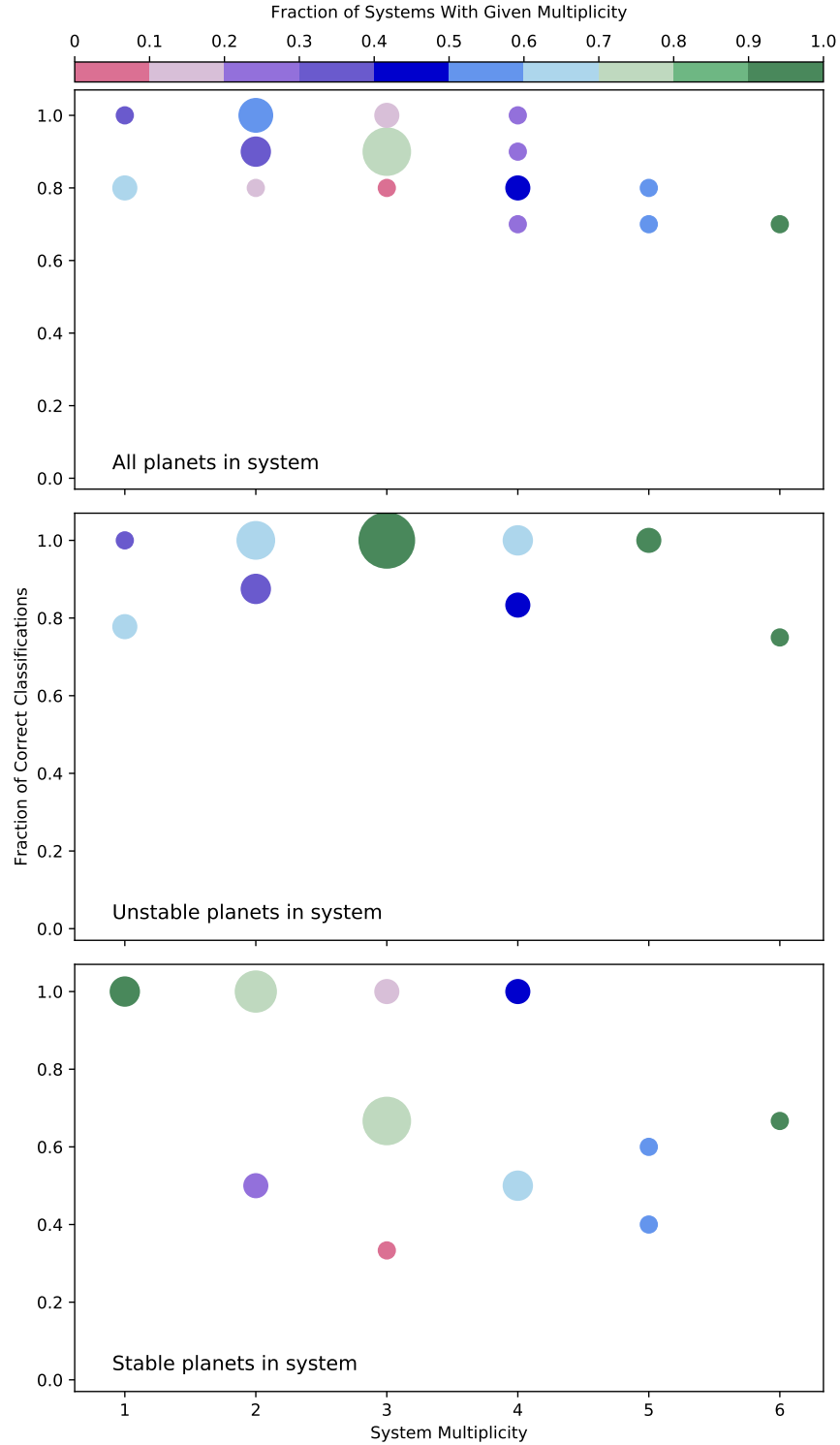
Figure 3.5: Final system multiplicity vs the fraction of correct classifications per system. The color of a point represents the fraction of systems with a given multiplicity while size of the point scales with the total number of systems. The panels, from top to bottom, show the fraction of all planets in a system correctly classified, the fraction of unstable planets in a system correctly classified, and the fraction of stable planets in a system correctly classified. My classifier performs very well on unstable planets but has a lower accuracy on high multiplicity systems.

These systems are (as anticipated) outliers from the majority of other systems in the ensemble. The eight planet system contains no Jovian planets, and the two most massive planets are ice giants; the remaining planets in the system are Earths or super Earths. In the two six planet systems, both have six or more super Earths, and the high mass planets are mostly Neptune mass planets. One has a single Jupiter mass planet. All of these systems exhibit substantial planet-planet scattering over a majority of the 10 Myr evolution, but few (if any) of the planets are massive enough to cause scattering strong enough to cause ejection. As such, collisions dominate the planet loss mechanisms in these systems: the eight planet system loses both unstable planets to collisions and the six planet systems each lose three of four unstable planets to collisions. The classifier does not have enough examples of the orbital evolution of such systems to perform an accurate classification.

4 Discussion

In this section, I investigate the major reasons for misclassification by my classifier and then compare the results of this work with other similar investigations.

4.1 Reasons for Misclassification

There are two major reasons that my classifier fails to correctly predict the fate of planets in the simulations: rarity and time scale.

Rarity is the lack of many representative examples in the simulation ensemble. Rarity is the cause of most of the MUS misclassifications. Of the 10 total misclassified unstable planets, seven were lost to collisions. Instability due to collision accounts for only 22 percent of planets lost in the training set, and collisions are often challenging to identify before they happen. A collision can occur at any time to planets of any mass, and they can happen in many different orbital states, such as when a planet is already undergoing some sort of catastrophic instability (i.e., a planet already ejecting from the system) or those just beginning to scatter. Due to the relative inhomogeneity of planets suffering collisions, there is no statistically strong basis for the classifier to learn what a collision looks like. I could potentially increase the statistical significance of collisional examples with more simulations in the ensemble.

The other major reason for misclassification of planets in the testing set is timescale. Because my classifier uses a only small fraction of a planet's orbital history, it is blind to late-time orbital interactions or dynamical evolution caused through a chain of multiple interactions. Figure 4.1 shows the fraction of planets
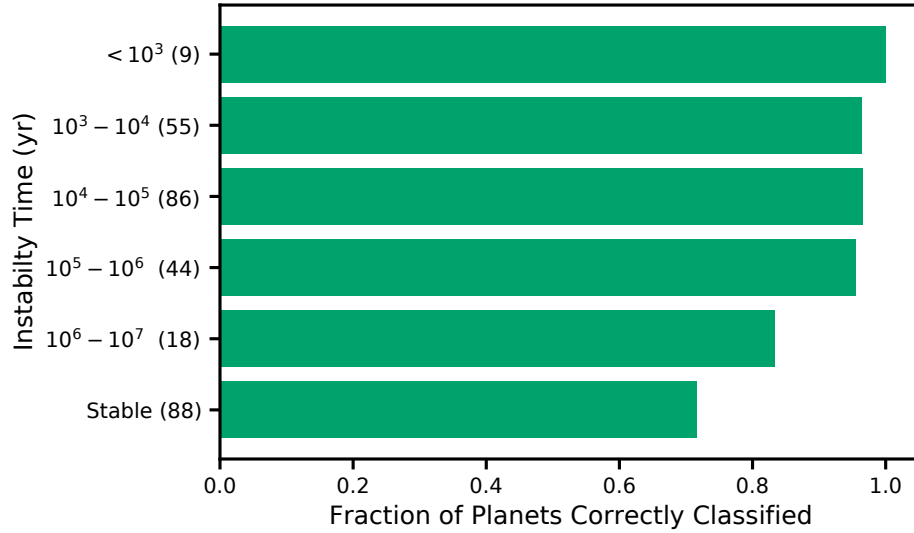
Figure 4.1: Classification accuracy as a function of instability time. Each bar shows the fraction of planets in a bin that were correctly classified as a function of instability time. The classifier achieves high accuracy in identifying planets that suffer catastrophic instability in the first 10^5 years but performs worse on objects that have interactions at later times.

correctly classified as a function of instability time. For short to intermediate instability times ($0 - 10^5$ yr), I have very high classification accuracy. Planets that suffer instability on these timescales likely have some indicator of future orbital evolution in their features: either they are too close to a massive planet, or their orbit is undergoing change on short timescales. I see a substantial decrease in classification accuracy with increasing instability time. Only 72 percent of stable planets are correctly identified by my classifier.

Examples of this timescale problem are shown for both stable and unstable planets in Figure 4.2. On the time scale relevant to the classifier (137 years), the MSU planet appears to be nearly static in a and e. It is not until a close interaction with several other planets at about 0.3 Myr that the MUS planet suffers large variations in eccentricity and semi-major axis. Conversely, the two MUS planets display visible orbital change in the first few hundred years. A late time interaction after 1 Myr causes their orbits finally settle into a stable configuration.

The issue of time scale will persist with any machine learning method that utilizes a subset of a simulation to predict instability. Due to the chaotic nature of orbital interactions, there is no easy way to predict Myr to Gyr orbital interactions, especially in my highly diverse planetary systems.

4.2   Comparison With Other Works

In this paper, I have presented a machine learning classifier designed to predict instability that has been trained on highly diverse planetary systems (in terms of mass and dynamical spacing). I now compare the performance of my classifier against the performance of other similar methods. Tamayo et al. (2016) and Tamayo et al. (2020) develop gradient boosting classifiers to predict the dissolution of compact three
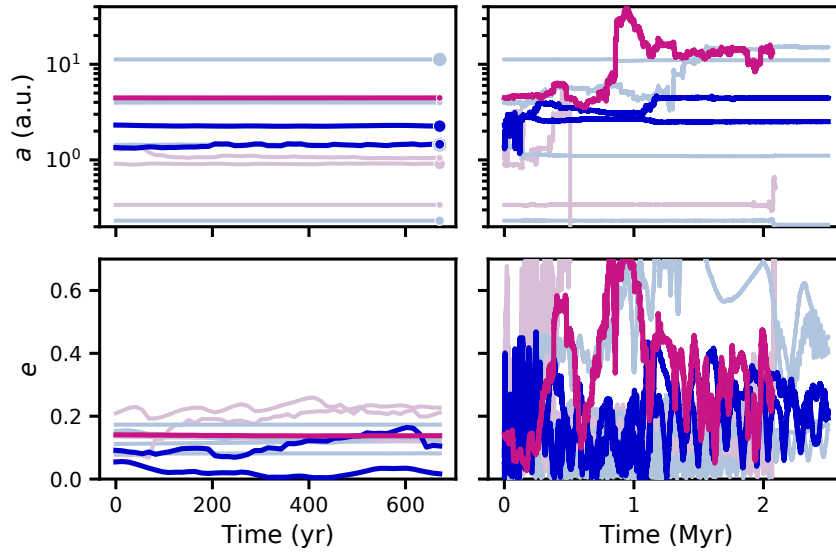
Figure 4.2: Orbital evolution of one system in time. The left panels show the orbits of the ten planets for the first few hundred years of the simulation (a similar scale as that used for classification), while the right panels show 2.5 Myr of orbital history. The top row depicts semi-major axis evolution and the bottom row shows eccentricity. The small points on the right side of the upper left panel are scaled to show relative planet masses. Coloring conventions follow Figure 3.2. Late time interactions that are unknowable to the classifier can cause planets in the testing set to be misclassified.

planet systems. Tamayo et al. (2020) compares the SPOCK classifier from that work with simple classifiers derived from constraints from MENGO (Cincotta et al., 2003), AMD (Laskar, 2000), and the Hill criterion (Chambers et al., 1996). I can utilize the 'area under the curve' (AUC) metric, which is a measure of relative model performance on the given data, to compare with the performance of three methods presented in Tamayo et al. (2020). A perfect model would have an AUC of 1 and a model that returns a random value (no informed classification) would have an AUC of 0.5. Tamayo et al. (2020) report that SPOCK has an AUC of 0.98, MENGO has an AUC of 0.95, and Hill has an AUC of 0.82. Comparatively, my classifier has an AUC of 0.91. Given the substantial differences in the types of systems being classified in these works, the fiducial classifier presented in this work performs quite well when compared to other models in the field.

Tamayo et al. (2020) and Cranmer et al. (2021) present a method of identifying stability in systems with multiplicities higher than three by classifying all three planet sub-systems in the high multiplicity system using their model (which has been trained on three planet systems). The general idea, inspired by Chambers et al. (1996), states that neighboring planets have the most impact on a given planet's stability. Because I have classified systems of ten planets using the dynamical influence of the most massive planets as the important features in determining stability, I ask the reverse of the Tamayo et al. (2020) question: is the influence of neighboring planets (effectively, the three planet sub-system) important for stability in my analysis? Figure 4.3 shows all of the 3-planet sub-systems in the 30 ten planet systems in my testing set. I separate them into categories based on the number of the three most massive planets that exist in the sub-system. If the neighboring planets were the most important for determining stability, I would expect that sub-systems without any of the most massive planets to have substantially fewer planets correctly classified.
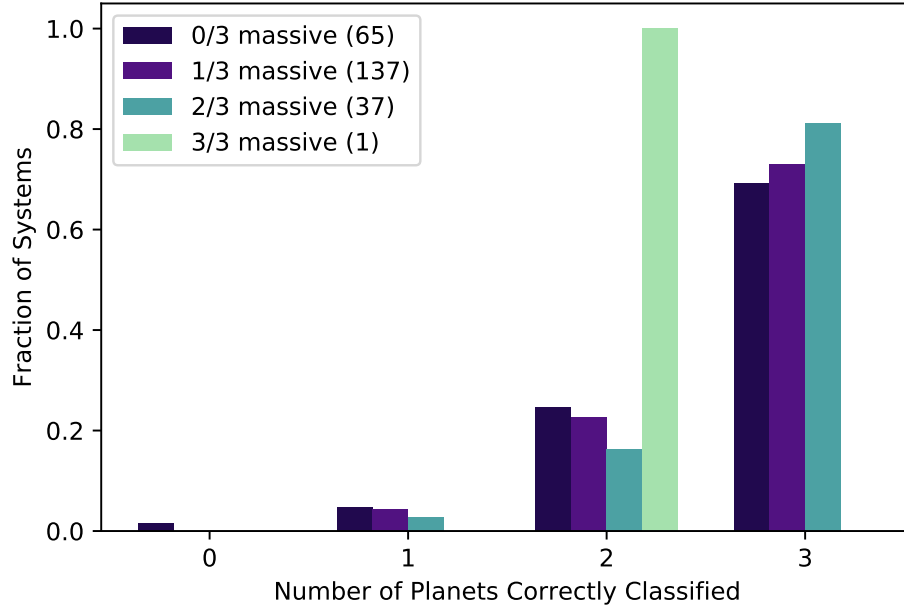
Figure 4.3: Classification accuracy for three planet sub-systems. Each bar color denotes three planet sub-systems that contain none of the three most massive planets in the parent 10 planet system (dark blue), one of the three most massive planets (purple), two of the three most massive planets (teal) or all three of the most massive planets (light green). The horizontal axis shows the number of planets correctly classified in the three planet system. There is little dependence in classification accuracy on the presence of a neighboring massive planet.

Instead, I see that the fraction of planets correctly classified is roughly constant (within statistical uncertainty) regardless of the number of neighbors that were included in the feature set. This suggests that, for systems with a substantial planet mass disparity (such as the Solar System or the planetary systems used in this work), the most massive planets play an important role in determining the stability of all planets in the system.

5 Conclusion

In this work, I use numerical simulations of diverse planetary systems to train a random forest machine learning classifier to predict the stability of individual planets. I derive dynamically motivated features from the simulations based on a 137 year window. I find the following:

1. Using only twelve features based on a planet's orbit and its relationship to the three most massive planets in the system, I am able to identify unstable planets with a > 88 percent accuracy. The most predictive power comes from the mass ratios with the three most massive planets in the system and from the standard deviation of a planet's eccentricity. My classifier tends to assign higher probabilities

to planets that are correctly classified.

2. Misclassified planets tend to lie along boundaries in feature space, indicating that their orbital evolution is somewhat ambiguous. The classifier performs better on unstable planets (95 percent accuracy) than stable planets (72 percent accuracy) due to the larger statistical sample in the planet ensemble.

3. I find that the major reason for misclassification by the classifier is the short time window upon which I compute features. Due to continuing orbital interactions between the ten planets of varied mass in these systems, the classifier has no way of predicting chaotic interactions at late times. Other misclassifications arise due to the rarity of interactions like collisions in the ensemble; the classifier does not have sufficient examples to create a robust representation of these events. Some further accuracy could likely be gained with a larger number of systems in both the training and testing sets, but time scales will always be a fundamental limit on the accuracy of machine learning methods such as used herein.

Machine learning has been shown to be a valuable addition to investigations of planetary stability with numerical simulations. Machine learning methods allow for substantially reduced computational cost compared to numerical simulations, without sacrificing as much accuracy as traditional metrics for stability characterization.

Future Work

This work serves as another contribution to the growing field machine learning and computing techniques applied to orbital dynamics, and more broadly, astrophysics. However, the methods used in this work do pose some limitations that could be addressed in subsequent research.

For one, the 10^7 year integration timescale is much shorter than the average lifespan of known planetary systems. Exploring the evolution of these planetary systems for longer timescales may change the performance and interpretation of my classifier. Also important to note is the limitations of the Mordasini planet population itself. This planet population is only one of many proposed initial distributions of planets; future work could further explore the performance of this method on different parameterizations of initial conditions.

In this paper, I present a 2-class classifier, where any given planet is predicted to be either stable or unstable. However, the nature of orbital fate offers the potential to expand classification capabilities. Beyond the stable-unstable boundary, I could explore a classifier's predictive power to distinguish between unstable and scattering planets. Similarly, this methodology could also be applied to predict specific planet fates

(predicting ejections, planet-planet collisions, planet-central body collisions, and surviving planets), rather than just long-term stability or instability.

While my simulations use exclusively single-star systems, of the exoplanetary systems discovered in the past decade, nearly 60 percent are circumbinary systems (with two stars). The applications of machine learning to predicting stable configurations in multi-planet circumbinary systems is fairly unexplored (e.g, Lam & Kipping, 2018), but could be extended to study a wider diversity of systems as presented in this work.

Broader Context and Implications

The general problem of planetary stability–especially in such diverse, packed systems–is complicated, chaotic, and computationally prohibitive. The applications of machine learning to orbital dynamics set the precedent for a fundamentally new way to explore planetary dynamics. There exists a computational bottleneck in this realm of research, but these machine learning methods drastically accelerate the process to analytically predict unstable configurations. While this work doesn't solve the problem of generic planetary stability, I can reliably identify instability in these highly diverse systems in a computationally efficient manner.

Thus far, the exoplanets discovered have been within a relatively small region of physical space. Yet, already, nearly half of these confirmed exoplanets have been found in multi-planet systems. As the number of observed multi-planet exoplanetary systems grows, it becomes less practical to study each one in detail. Astronomers will inevitably discover new exoplanetary systems with different orbital configurations; therefore, methods like those presented in this paper will allow us to understand the orbital evolution and stability of these diverse, multi-planet systems. More broadly, with an understanding of stability of planetary systems, this research can help inform the range of planet compositions, configurations, and system architectures that exist in exoplanetary systems.

With a high accuracy and minimized computational power, methods like those presented in this research offer potential to analyze planetary stability more precisely and on a larger scale than ever before. The insight gained on the dynamics of these diverse multi-planet systems could also lead to discoveries about the formation of planetary systems, as well as their past and future evolution. Not only is this applicable to the new exoplanetary systems we discover, but also within our own Solar System. With this knowledge of planets' formation, evolution, and the underlying physics of their orbital dynamics, we can begin to better understand our role in the universe.

Bibliography

Agnew M. T., Maddison S. T., Horner J., Kane S. R., 2019, *Predicting multiple planet stability and habitable zone companions in the TESS era*, Monthly Notices of the Royal Astronomical Society, 485, 4703

Barnes R., Quinn T., 2004, *The (In)stability of Planetary Systems*, The Astrophysical Journal, 611, 494

Carruba V., Aljbaae S., Domingos R. C., Lucchini A., Furlaneto P., 2020, *Machine learning classification of new asteroid families members*, Monthly Notices of the Royal Astronomical Society

Chambers J. E., Migliorini F., 1997, *Mercury - A New Software Package for Orbital Integrations*, American Astronomical Society, 29

Chambers J. E., Wetherill G., Boss A. P., 1996, *The Stability of Multi-Planet Systems*, Icarus, 119, 261

Chatterjee S., Ford E. B., Matsumura S., Rasio F. A., 2008, *Dynamical Outcomes of Planet-Planet Scattering*, The Astrophysical Journal, 686, 580

Cincotta P. M., Giordano C. M., Simó C., 2003, *Phase space structure of multi-dimensional systems by means of the mean exponential growth factor of nearby orbits*, Physica D: Nonlinear Phenomena, 182, 151

Cranmer M., Tamayo D., Rein H., Battaglia P., Hadden S., Armitage P. J., Ho S., Spergel D. N., 2021, *A Bayesian neural network predicts the dissolution of compact planetary systems*, arXiv:2101.04117 [astro-ph, stat]

Deck K. M., Payne M., Holman M. J., 2013, *First-Order Resonance Overlap and the Stability of Close Two-Planet Systems*, ApJ, 774, 129

Faber P., Quillen A. C., 2007, *The total number of giant planets in debris discs with central clearings*, Monthly Notices of the Royal Astronomical Society, 382, 1823

Fang J., Margot J.-L., 2013, *Are Planetary Systems Filled to Capacity? A Study Based on Kepler Results*, The Astrophysical Journal, 767, 115

Gladman B., 1993, *Dynamics of Systems of Two Close Planets*, Icarus, 106, 247

Gratia P., Lissauer J. J., 2021, *Eccentricities and the stability of closely-spaced five-planet systems*, Icarus, 358, 114038

Jurić M., Tremaine S., 2008, *Dynamical Origin of Extrasolar Planet Eccentricity Distribution*, The Astrophysical Journal, 686, 603

Kratter K. M., Shannon A., 2014, *Planet packing in circumbinary systems*, Monthly Notices of the Royal Astronomical Society, 437, 3727

Lam C., Kipping D., 2018, *A Machine Learns to Predict the Stability of Circumbinary Planets*, Monthly Notices of the Royal Astronomical Society, 476, 5692

Laskar J., 2000, *On the Spacing of Planetary Systems*, Physical Review Letters, 84, 3240

Lissauer J. J., Gavino S., 2021, *Orbital stability of compact three-planet systems, I: Dependence of system lifetimes on initial orbital separations and longitudes*, Icarus, 364, 114470

Lissauer J. J., et al., 2011a, *Architecture and Dynamics of Kepler's Candidate Multiple Transiting Planet Systems*, The Astrophysical Journal Supplement Series, 197, 8

Lissauer J. J., et al., 2011b, *A closely packed system of low-mass, low-density planets transiting Kepler-11*, Nature, 470, 53

Mayor M., Queloz D., 1995, *A Jupiter-mass companion to a solar-type star*, Nature, 378, 355

Mordasini C., Alibert Y., Benz W., 2009a, *Extrasolar planet population synthesis I. Method, formation tracks, and mass-distance distribution*, Astronomy and Astrophysics, 501, 1139

Mordasini C., Alibert Y., Benz W., Naef D., 2009b, *Extrasolar planet population synthesis II: Statistical comparison with observation*, Astronomy & Astrophysics, 501, 1161

Morrison S. J., Kratter K. M., 2016, *Orbital stability of multi-planet systems: behavior at high masses*, The Astrophysical Journal, 823, 118

Obertas A., Van Laerhoven C., Tamayo D., 2017, *The stability of tightly-packed, evenly-spaced systems of Earth-mass planets orbiting a Sun-like star*, Icarus, 293, 52

Pu B., Wu Y., 2015, *Spacing of Kepler Planets: Sculpting by Dynamical Instability*, The Astrophysical Journal, 807, 44

Raymond S. N., Armitage P. J., Gorelick N., 2010, *Planet-Planet Scattering in Planetesimal Disks. II. Predictions for Outer Extrasolar Planetary Systems*, The Astrophysical Journal, 711, 772

Shikita B., Koyama H., Yamada S., 2010, *The Dynamics of Three-Planet Systems: An Approach from a Dynamical System*, The Astrophysical Journal, 712, 819

Smith A. W., Lissauer J. J., 2009, *Orbital stability of systems of closely-spaced planets*, Icarus, 201, 381

Smullen R. A., Volk K., 2020, *Machine Learning Classification of Kuiper Belt Populations*, Monthly Notices of the Royal Astronomical Society, 497, 1391

Smullen R. A., Kratter K. M., Shannon A., 2016, *Planet Scattering Around Binaries: Ejections, Not Collisions*, Monthly Notices of the Royal Astronomical Society, 461, 1288

Tamayo D., et al., 2016, *A Machine Learns to Predict the Stability of Tightly Packed Planetary Systems*, The Astrophysical Journal, 832, L22

Tamayo D., et al., 2020, *Predicting the long-term stability of compact multiplanet systems*, Proceedings of the National Academy of Sciences, 117, 18194

Wisdom J., 1980, *The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body problem*, The Astronomical Journal, 85, 1122

Wu D.-H., Zhang R. C., Zhou J.-L., Steffen J. H., 2019, *Dynamical instability and its implications for planetary system architecture*, Monthly Notices of the Royal Astronomical Society, 484, 1538